



(19)

(11) Publication number: **200**

Generated Document.

**PATENT ABSTRACTS OF JAPAN**(21) Application number: **2003085671**(51) Intl. Cl.: **G06F 3/06**(22) Application date: **08.12.00**

(30) Priority:	(71) Applicant: <b>TOSHIBA CORP</b>
(43) Date of application publication: <b>24.10.03</b>	(72) Inventor: <b>SASAMOTO KYOICHI TAKAKUWA MASAYU</b>
(84) Designated contracting states:	(74) Representative:

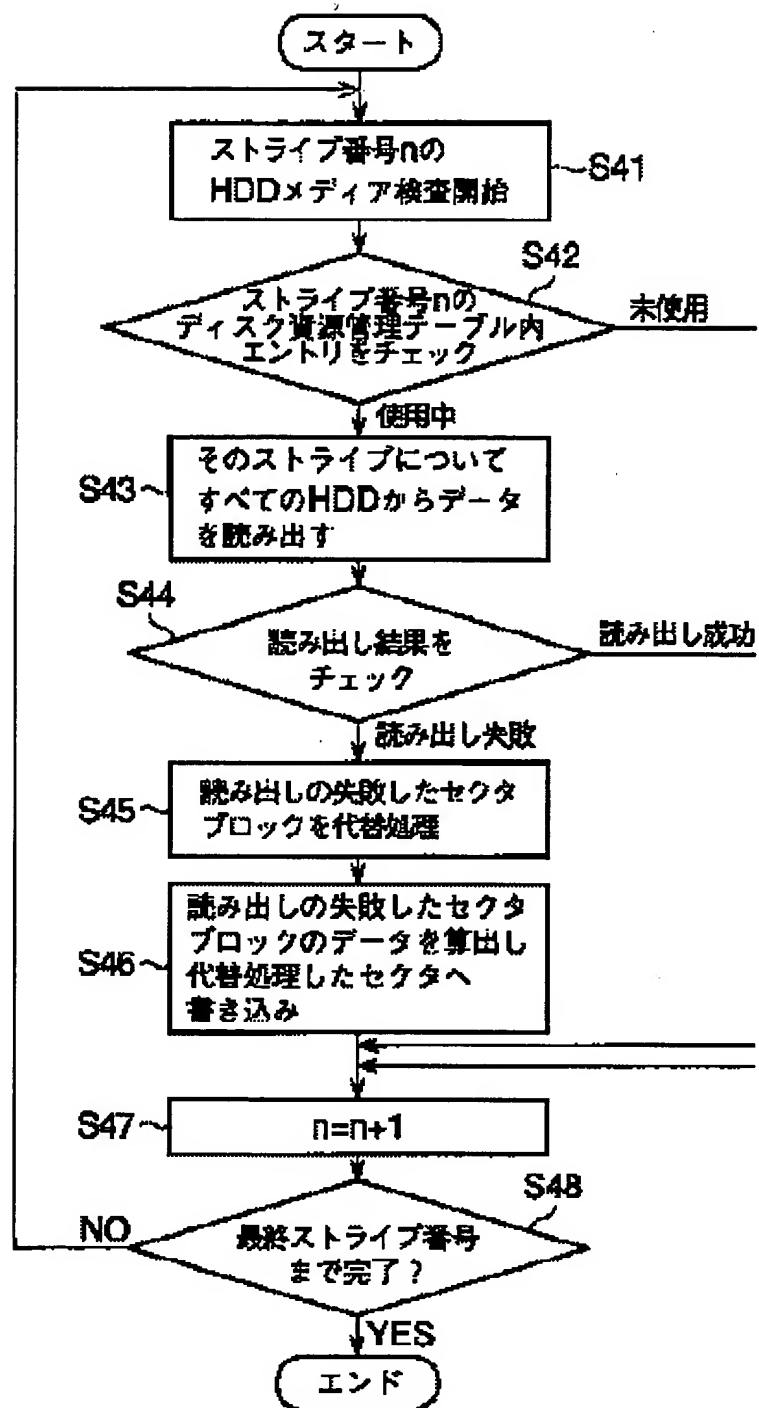
**(54) METHOD FOR DATA  
RECOVERY AND DISK  
ARRAY CONTROLLER IN  
DISK ARRAY APPARATUS**

(57) Abstract:

**PROBLEM TO BE SOLVED:** To achieve quick detection and recovery of media failures by executing media check processing with a distinction between areas actually in use and not in use at a file system among disk areas of a disk array.

**SOLUTION:** In a media check processing for checking partial failures of a plurality of HDDs (hard disk drives) composed of the disk array, it is determined whether each stripe in the disk areas of the disk array is used by the file system or not based on a disk source management table (S41, S42), and a media check including data reading from the HDDs is executed only for a stripe (a first stripe) used by the file system (S43, S44).

COPYRIGHT: (C)2004,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2003-303057  
(P2003-303057A)

(43) 公開日 平成15年10月24日 (2003. 10. 24)

(51) Int.Cl.<sup>7</sup>  
G 0 6 F 3/06

識別記号  
3 0 6  
5 4 0

F I  
G 0 6 F 3/06

データコード\* (参考)

3 0 6 K 5 B 0 6 5  
5 4 0

審査請求 未請求 請求項の数 4 O L (全 16 頁)

(21) 出願番号 特願2003-85671(P2003-85671)  
(62) 分割の表示 特願2000-374616(P2000-374616)の  
分割  
(22) 出願日 平成12年12月8日 (2000. 12. 8)

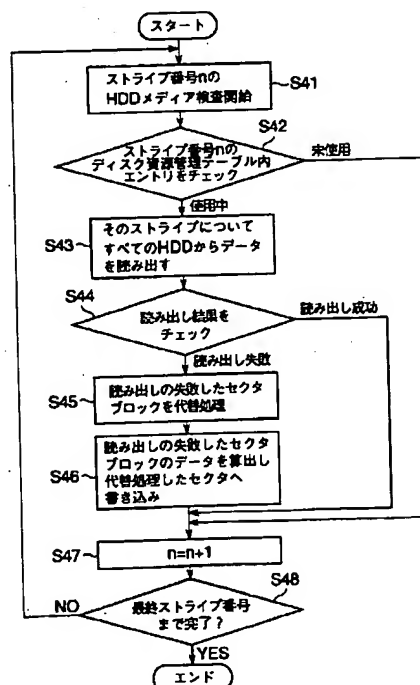
(71) 出願人 000003078  
株式会社東芝  
東京都港区芝浦一丁目1番1号  
(72) 発明者 笹本 享一  
東京都府中市東芝町1番地 株式会社東芝  
府中事業所内  
(72) 発明者 高桑 正幸  
東京都府中市東芝町1番地 株式会社東芝  
府中事業所内  
(74) 代理人 100058479  
弁理士 鈴江 武彦 (外5名)  
Fターム(参考) 5B065 BA01 CA30 EA18

(54) 【発明の名称】 ディスクアレイ装置におけるデータ修復方法及びディスクアレイコントローラ

(57) 【要約】

【課題】 ディスクアレイのディスク領域のうちファイルシステムにて実際に使用されている領域と使用されていない領域とを区別してメディア検査処理を実行することで、メディア障害の早期検出・早期修復を実現する。

【解決手段】 ディスクアレイを構成する複数のHDDの部分的な障害を検出するメディア検査処理において、ディスクアレイのディスク領域の各ストライプについて、ディスク資源管理テーブルに基づいて、そのストライプがファイルシステムにより使用されているか否かを判定し (S41, S42)、ファイルシステムにより使用されているストライプ (第1のストライプ) についてのみ、ディスクドライブからのデータ読み出しを含むメディア検査を実行する (S43, S44)



## 【特許請求の範囲】

【請求項1】 複数のディスクドライブから構成されるRAID (Redundant Arrays of Inexpensive Disks) 構成のディスクアレイを備えたディスクアレイ装置におけるデータ修復方法であって、  
 前記ディスクアレイのディスク領域を論理ブロック単位に管理するファイルシステムを備えたホスト計算機から、前記ファイルシステムにより使用されている論理ブロックまたは当該論理ブロックを含むストライプを示す第1のディスク資源管理情報を取得するステップと、  
 前記第1のディスク資源管理情報から前記ディスクアレイのディスク領域内の各ストライプ毎に前記ファイルシステムにより使用されている論理ブロックを含むか否かを示す第2のディスク資源管理情報を生成して前記ディスクアレイ装置内に保持するステップと、  
 前記ホスト計算機からデータ書き込み要求を受け取った場合、当該要求で指定されたデータ書き込み先が属するストライプを特定するステップと、  
 前記特定されたストライプが前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、前記ディスクアレイ装置内に保持されている前記第2のディスク資源管理情報に基づいて判定するステップと、  
 前記特定されたストライプが前記第2のストライプであると判定された場合、当該特定されたストライプが前記第1のストライプであることを示すように前記ディスクアレイ装置内に保持されている前記第2のディスク資源管理情報を更新するステップと、  
 前記ディスクアレイ内の前記各ディスクドライブの記憶内容を読み出すことにより当該ディスクドライブの部分的な障害を検出するメディア検査処理を実行する場合に、前記ディスクアレイのディスク領域のすべてのストライプについて、前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか、或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、前記第2のディスク資源管理情報に基づいて判定するステップと、  
 前記判定ステップでの判定結果をもとに、前記第1のストライプについてのみ、前記ディスクドライブからのデータ読み出しを含むメディア検査を実行するステップと、  
 前記メディア検査で障害が検出された箇所のデータをRAID機能により修復するステップとを具備することを特徴とするディスクアレイ装置におけるデータ修復方法。

【請求項2】 複数のディスクドライブから構成されるRAID (Redundant Arrays of Inexpensive Disks)

構成のディスクアレイを備えたディスクアレイ装置におけるデータ修復方法であって、  
 前記ディスクアレイのディスク領域を論理ブロック単位に管理するファイルシステムを備えたホスト計算機から、前記ファイルシステムにより使用されている論理ブロックまたは当該論理ブロックを含むストライプを示す第1のディスク資源管理情報を取得するステップと、  
 前記第1のディスク資源管理情報から前記ディスクアレイのディスク領域内の各ストライプ毎に前記ファイルシステムにより使用されている論理ブロックを含むか否かを示す第2のディスク資源管理情報を生成して前記ディスクアレイ装置内に保持するステップと、  
 前記ホスト計算機からデータ書き込み要求を受け取った場合、当該要求で指定されたデータ書き込み先が属するストライプを特定するステップと、  
 前記特定されたストライプが前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、前記ディスクアレイ装置内に保持されている前記第2のディスク資源管理情報に基づいて判定するステップと、  
 前記特定されたストライプが前記第2のストライプであると判定された場合、当該特定されたストライプが前記第1のストライプであることを示すように前記ディスクアレイ装置内に保持されている前記第2のディスク資源管理情報を更新するステップと、  
 前記ディスクアレイ内の前記各ディスクドライブの記憶内容を読み出すことにより当該ディスクドライブの部分的な障害を検出するメディア検査処理を前記ディスクアレイのディスク領域内の全てのストライプについてストライプ毎に順次実行するステップと、  
 前記メディア検査処理で前記ディスクドライブの部分的な障害が検出された場合、当該障害が検出された箇所を含むストライプが前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか、或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、前記ディスクアレイ装置内に保持されている前記第2のディスク資源管理情報に基づいて判定するステップと、  
 前記判定ステップで前記第1のストライプであると判定され、且つ前記ディスクアレイ内の前記複数のディスクドライブのうち前記障害が検出された前記ディスクドライブ以外のディスクドライブが正常な場合、前記障害が検出された箇所のデータをRAID機能により修復するステップと、  
 前記判定ステップで前記第2のストライプであると判定された場合、前記障害が検出された箇所のデータを固定データにより修復するステップとを具備することを特徴とするディスクアレイ装置におけるデータ修復方法。

【請求項3】 複数のディスクドライブから構成されるRAID (Redundant Arrays of Inexpensive Disks) 構成のディスクアレイを制御するディスクアレイコントローラにおいて、

前記ディスクアレイのディスク領域を論理ブロック単位に管理するファイルシステムを備えたホスト計算機から送信される、前記ファイルシステムにより使用されている論理ブロックまたは当該論理ブロックを含むストライプを示す第1のディスク資源管理情報から、前記ディスクアレイのディスク領域内の各ストライプ毎に前記ファイルシステムにより使用されている論理ブロックを含むか否かを示す第2のディスク資源管理情報を生成する手段と、

前記第2のディスク資源管理情報が格納されるメモリと、

前記ホスト計算機からデータ書き込み要求を受け取った場合、当該要求で指定されたデータ書き込み先が属するストライプを特定する手段と、

前記特定手段により前記データ書き込み先が属するストライプが特定された場合に、当該ストライプが、前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか、或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを判定する第1の判定手段と、

前記特定手段により特定されたストライプが前記第2のストライプであると前記第1の判定手段により判定された場合、当該ストライプが前記第1のストライプであることを示すように前記メモリに格納されている前記第2のディスク資源管理情報を更新する手段と、

前記ディスクアレイのディスク領域のすべてのストライプについて、前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか、或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、前記第2のディスク資源管理情報に基づいて判定する第2の判定手段と、

前記ディスクアレイ内の前記各ディスクドライブの記憶内容を読み出すことにより当該ディスクドライブの部分的な障害を検出するメディア検査手段であって、前記第2の判定手段により前記第1のストライプであると判定されたストライプだけを対象に、前記ディスクドライブからのデータ読み出しを含むメディア検査を実行するメディア検査手段と、

前記メディア検査手段により障害が検出された箇所のデータをRAID機能により修復するデータ修復手段とを具備することを特徴とするディスクアレイコントローラ。

【請求項4】 複数のディスクドライブから構成されるRAID (Redundant Arrays of Inexpensive Disks) 構成のディスクアレイを制御するディスクアレイコント

ローラにおいて、

前記ディスクアレイのディスク領域を論理ブロック単位に管理するファイルシステムを備えたホスト計算機から送信される、前記ファイルシステムにより使用されている論理ブロックまたは当該論理ブロックを含むストライプを示す第1のディスク資源管理情報から、前記ディスクアレイのディスク領域内の各ストライプ毎に前記ファイルシステムにより使用されている論理ブロックを含むか否かを示す第2のディスク資源管理情報を生成する手段と、

前記第2のディスク資源管理情報が格納されるメモリと、

と、

前記ホスト計算機からデータ書き込み要求を受け取った場合、当該要求で指定されたデータ書き込み先が属するストライプを特定する手段と、

前記特定手段により前記データ書き込み先が属するストライプが特定された場合に、当該ストライプが、前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか、或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを判定する第1の判定手段と、

前記特定手段により特定されたストライプが前記第2のストライプであると前記第1の判定手段により判定された場合、当該ストライプが前記第1のストライプであることを示すように前記メモリに格納されている前記第2のディスク資源管理情報を更新する手段と、

前記ディスクアレイ内の前記各ディスクドライブの記憶内容を読み出すことにより当該ディスクドライブの部分的な障害を検出するメディア検査処理を前記ストライプ毎に実行するメディア検査手段と、

前記メディア検査手段により前記ディスクドライブの部分的な障害が検出された場合、当該障害が検出された箇所を含むストライプが前記ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか、或いは前記ファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、前記第2のディスク資源管理情報に基づいて判定する第2の判定手段と、

前記第2の判定手段により前記第1のストライプであると判定され、且つ前記ディスクアレイ内の前記複数のディスクドライブのうち前記障害が検出された前記ディスクドライブ以外のディスクドライブが正常な場合、前記障害が検出された箇所のデータをRAID機能により修復する第1のデータ修復手段と、

前記第2の判定手段により前記第2のストライプであると判定された場合、前記障害が検出された箇所のデータを固定データにより修復する第2のデータ修復手段とを具備することを特徴とするディスクアレイコントローラ。

【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、RAID (Redundant Arrays of Inexpensive Disks) 構成のディスクアレイ装置におけるデータ修復方法に係り、特にディスクアレイを構成するメンバーのディスクドライブの部分的な障害（メディア障害）を検出して修復する場合に好適なデータ修復方法及びディスクアレイコントローラに関する。

## 【0002】

【従来の技術】一般にディスクアレイ装置は、複数のディスクドライブ、問えば磁気ディスク装置（以下、HDDと称する）から構成されるディスクアレイと、このディスクアレイ内の各HDD（メンバーHDD）に対するアクセスを制御するディスクアレイコントローラとを備え、当該各HDDを並列に動かして読み出し／書き込みを分散して実行することでアクセスの高速化を図ると共に、冗長構成によって信頼性の向上を図るようにした外部記憶装置として知られている。

【0003】上記ディスクアレイコントローラは、ホスト計算機から転送される書き込みデータに対して、データ訂正情報としての冗長データを生成し、上記複数のHDDのうちのいずれかに書き込むようになっている。これにより、複数のHDDのうちの1台の故障に対し、この冗長データと残りのHDDのデータを用いて故障したHDDのデータを修復することを可能としている。

【0004】データ冗長化の手法の1つとして、RAIDの手法が知られている。RAID手法では、RAIDのデータと冗長データとの関連において、種々のRAIDレベルに分類されている。RAIDレベルの代表的なものにレベル3とレベル5がある。

【0005】レベル3（RAIDレベル3）では、ホスト計算機から転送される更新データ（書き込みデータ）を分割して、その分割された更新データ間の排他的論理和演算を行うことで冗長データとしてのパリティデータを生成し、当該パリティデータで複数のHDDのいずれかに書き込まれている元のパリティデータを更新する。一方、レベル5（RAIDレベル5）では、ホスト計算機から転送される更新データ（新データ）と、当該更新データの格納先となるHDD内領域に格納されている更新前のデータ（旧データ）と、当該更新データの格納先に対応する別のHDDの領域に格納されている更新前のパリティデータ（旧パリティデータ）との間の排他的論理和演算を行うことで、更新されたパリティデータ（新パリティデータ）を生成し、当該更新パリティデータで元のパリティデータを更新する。

【0006】このようなRAID構成のディスクアレイ装置では、ディスクアレイ内のメンバーHDDが故障した場合に、故障したHDD以外のHDDから、ディスクアレイのディスク領域を管理する単位であるストライプ毎にデータを読み出して、それらのデータの排他的論理

和演算を行うRAIDの機能により、故障したHDDのすべての領域のデータを、故障したHDDに代えて用いられるHDD内に修復することができる。この故障したHDDに代えて用いられるHDDは、故障したHDDと交換して用いられるHDD、またはディスクコントローラに予め接続されていて、故障したHDDの代替として割り付けられるスペアHDDである。

## 【0007】

【発明が解決しようとする課題】このように、RAID構成のディスクアレイ装置では、HDDが故障しても、故障したHDDのデータを元通りに修復することができる。

【0008】ところが従来のディスクアレイ装置では、故障したHDD内のデータを元通りに修復するのに、すべてのHDD領域のデータをRAIDの機能により修復していた。このため、近年のようにHDD容量が増加するに伴い、データの修復に非常に時間がかかるという問題があった。

【0009】また、データ修復をしている最中は、一般にRAIDによるデータの冗長性が損なわれる。このため、修復に時間がかかるほど更に他のHDD故障も発生しやすくなって、データ修復不能となり、データが消失する危険性が高まる。また、故障したHDDのデータを修復するためには、その他のHDDの全領域を読み出す必要がある。もし、この読み出しの対象となるHDDにてメディア障害（HDDの部分的な障害）が発生すると、HDDの多重障害となってデータ修復不能となるため、ディスクアレイのディスク領域（ディスクボリューム）の閉塞またはデータ修復処理の継続不能となる。これによりディスクアレイ装置の信頼性が低下する。

【0010】ところが本発明者は、ディスクアレイ装置を利用するホスト計算機のファイルシステムが実際に使用しているHDD領域以外は、データ修復処理は必ずしも必要でないことを想到するに至った。そこで、ファイルシステムが実際に使用しているHDD領域（ディスク領域）のみを対象にHDD（ディスクドライブ）のデータ修復を行うならば、データ修復処理に要する時間を短縮して、HDD故障などの危険性を減らすことが可能となる。しかし、従来のディスクアレイ装置は、ホスト計算機のファイルシステムが実際に使用しているHDD領域を知ることができず、したがってファイルシステムが使用しているHDD領域のみを対象にHDDのデータ修復を行うことはできない。

【0011】また、HDDのメディア障害、つまりHDDの部分的な障害であるセクタブロックの障害を早期に検出しこれを修復する目的で、周期的にHDDの内容を読み出して検査するHDDメディア検査処理が一般に知られている。しかし従来のディスクアレイ装置では、ファイルシステムが使用しているHDD領域が不明のため、HDDの全領域について読み出し検査をする必要が

あった。このため、故障したHDDのデータの修復処理の場合と同様に、検査に非常に時間を要するという問題があった。

【0012】また、メディア障害を発見した際には、その障害箇所のセクタブロックを他のセクタブロックに代替する処理を行い、その代替先のブロック内に、代替元のデータを修復する必要がある。しかし、このデータ修復が行えるのは、RAIDの機能によりデータの冗長性が確保されている場合に限られる。したがって、HDD故障などに起因して行われるデータの修復中は、HDDメディア検査処理にてメディア障害を見つけても修復することができず、ディスクアレイ装置の信頼性が低下するという問題もあった。

【0013】本発明は上記事情を考慮してなされたものでその目的は、ディスクアレイのディスク領域のうちファイルシステムにて実際に使用されている領域と使用されていない領域とを区別してメディア検査処理を実行することで、メディア障害の早期検出・早期修復を実現し、これによりディスクアレイ装置の信頼性を向上できるようにすることにある。

【0014】

【課題を解決するための手段】本発明は、ディスクアレイのディスク領域を論理ブロック単位に管理するファイルシステムを備えたホスト計算機から、当該ファイルシステムにより使用されている論理ブロックまたは当該論理ブロックを含むストライプを示す第1のディスク資源管理情報を取得するステップと、上記第1のディスク資源管理情報から上記ディスク領域内の各ストライプ毎にファイルシステムにより使用されている論理ブロックを含むか否かを示す第2のディスク資源管理情報を生成してディスクアレイ装置内に保持するステップと、上記ホスト計算機からデータ書き込み要求を受け取った場合、当該要求で指定されたデータ書き込み先が属するストライプを特定するステップと、この特定されたストライプがファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか或いはファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、上記ディスクアレイ装置内に保持されている第2のディスク資源管理情報に基づいて判定するステップと、上記特定されたストライプが上記第2のストライプであると判定された場合、当該特定されたストライプが上記第1のストライプであることを示すように上記第2のディスク資源管理情報を更新するステップと、ディスクアレイ内の各ディスクドライブの記憶内容を読み出すことにより当該ディスクドライブの部分的な障害を検出するメディア検査処理を実行する場合に、ディスクアレイのディスク領域のすべてのストライプについて上記第1のストライプであるか、或いは上記第2のストライプであるかを上記第2のディスク資源管理情報に基づいて判定するステップと、第1のストライ

プ、即ちファイルシステムが使用しているストライプについてのみ、ディスクドライブからのデータ読み出しを含むメディア検査を実行するステップと、このメディア検査で障害が検出された箇所のデータをRAID機能により修復するステップとを備えたことを特徴とする。

【0015】このような構成においては、ディスクアレイ装置内の例えばディスクアレイコントローラにて、ホスト計算機から第1のディスク資源管理情報が取得され、当該情報から第2のディスク資源管理情報が生成されてディスクアレイ装置内の例えばディスクアレイコントローラに保持される。ここでのメディア検査処理は、ファイルシステムが実際に使用している論理ブロックを含んでいることが第2のディスク資源管理情報によって示されているストライプ（第2のストライプ）だけを対象に実行される。

【0016】このように、ファイルシステムにて使用されているストライプのみ抽出してメディア検査をすることにより、その検査に要する処理時間を短縮することができる。また、処理時間を短縮したことで、その結果としてメディア障害をより早期に見つけることが可能となり、ディスクドライブの信頼性向上が図れる。これによりディスクアレイ装置の信頼性も向上する。

【0017】ホスト計算機から取得した第1のディスク資源管理情報は、当該情報を取得した直前までの論理ブロックの使用状況を表す。したがって、それ以降に発生したファイルの更新（ディスクアレイ装置へのデータ書き込み）については反映されていない。しかし、ファイルの更新の度に最新の第1のディスク資源管理情報をホスト計算機から取得するのは大幅な性能低下となる。

【0018】そこで、ホスト計算機からのデータ書き込み要求をディスクアレイ装置が受け取った場合、つまりディスクアレイ装置でのファイル更新が発生する場合、当該データ書き込み要求に基づいて、ディスクアレイ装置にて保持している第2のディスク資源管理情報を、ストライプの最新の使用状況を表すようにディスクアレイ装置自身が更新するとよい。

【0019】また本発明は、上記メディア検査処理を全てのストライプについてストライプ毎に順次実行し、ディスクドライブの部分的な障害が検出された場合には、障害が検出された箇所を含むストライプが第1または第2のストライプのいずれであるかを判定し、第1のストライプの場合で、即ちファイルシステムが使用しているストライプである場合で、障害が検出されたディスクドライブ以外のディスクドライブが正常な場合には、障害が検出された箇所のデータをRAID機能により修復し、第2のストライプの場合で、即ちファイルシステムが使用していないストライプである場合には、障害が検出された箇所のデータを固定データにより修復することをも特徴とする。

【0020】このように、メディア検査で障害が検出さ

れた箇所を含むストライプがファイルシステムにより使用されている場合で、且つ障害が検出されたディスクドライブ以外のディスクドライブが正常な場合だけ、障害が検出された箇所のデータをRAID機能により修復する。これに対し、障害が検出された箇所を含むストライプがファイルシステムにより使用されていない場合には、そのストライプ内のデータを保持している必要はないため、障害が検出されたディスクドライブ以外のディスクドライブが正常であるか否かに無関係に、つまりRAID機能によるデータの冗長性が確保されているか否かに無関係に、障害が検出された箇所のデータを強制的に固定データにより修復することで、メディア障害の修復できる可能性が大幅に向上してディスクドライブの信頼性を向上し、これによりディスクアレ装置の信頼性も向上する。

【0021】なお、以上の方法に係る発明は、装置（ディスクアレコントローラ、または同ディスクアレコントローラを備えたディスクアレ装置、または同ディスクアレ装置を備えた計算機システム）に係る発明としても成立する。

【0022】

【発明の実施の形態】以下、本発明の実施の形態につき図面を参照して説明する。図1は本発明の一実施形態に係るディスクアレ装置を備えた計算機システムの構成を示すブロック図である。図1の計算機システムは、ホスト計算機10と、このホスト計算機10によって利用されるディスクアレ装置20とから構成される。

【0023】ホスト計算機10は、当該ホスト計算機10と接続されているディスクアレ装置20（のディスク領域）に格納されるファイルを管理するファイルシステム11を備えている。このファイルシステム11はOS（オペレーティングシステム）により提供される機能の一部である。

【0024】ホスト計算機10は、ディスクアレ装置20のディスク領域内のすべての論理ブロック（連続する複数の物理セクタブロックで構成される固定長のブロック）について、そのブロックがファイルシステム11により使用されている（つまり有効なデータが格納されている）か、或いは使用されていない（つまりデータは格納されておらず新しいデータを格納できる）かを示す管理テーブル（以下、ディスク資源管理テーブルと称する）12を、当該計算機10が持つ記憶装置、例えばHDD（図示せず）に保持している。この記憶装置がディスクアレ装置20であっても構わない。また、ディスク資源管理テーブル12が、ディスク領域内のすべての物理セクタブロックについて、そのブロックがファイルシステム11により使用されている否かを示すものであっても構わない。

【0025】ホスト計算機10（が持つ記憶装置）には、ファイルシステム11から予め定められたタイミン

グでディスク資源管理テーブル12を取得して、当該テーブル12から生成されるディスク資源管理情報リスト90（図9参照）をディスクアレ装置20に送信する専用ソフトウェア13がインストールされている。このディスク資源管理情報リスト90は、後述するようにファイルシステム11により使用されているすべての論理ブロックについて、そのブロックの識別情報としての論理ブロック番号の集合からなる。

【0026】ディスクアレ装置20は、ディスクアレ21とディスクアレコントローラ22とから構成されている。ディスクアレ21は、ディスクアレコントローラ22と接続される複数のディスクドライブ、例えば4台のHDD（磁気ディスク装置）210-0～210-3から構成される。ディスクアレコントローラ22には、HDD210-0～210-3のいずれかに障害が発生した場合のバックアップディスク用に割り当てられるスペアHDD（図示せず）も接続されている。

【0027】ディスクアレコントローラ22は、ディスクアレ21内の各HDD210-0～210-3に対するアクセスを制御する。ディスクアレコントローラ22は、当該コントローラ22の主制御部をなすマイクロプロセッサ221と、メモリ222とを備えている。メモリ222には、マイクロプロセッサ221が実行する制御プログラム222aが格納されている。またメモリ222には、ディスク資源管理テーブル領域222bが確保されている。このディスク資源管理テーブル領域222bは、ホスト計算機10から送信されるディスク資源管理情報リスト90をもとに生成されるディスク資源管理テーブル120を格納するのに用いられる。

【0028】本実施形態では、ディスクアレ装置20がRAID5レベルで用いられるものとする。この場合、HDD210-0～210-3がいずれもデータ並びにパリティデータ（冗長データ）の格納用（データ・パリティディスク用）に用いられる。なお、ディスクアレ装置20がRAID3レベルで用いられる場合には、HDD210-0～210-3のうちの3台がデータ格納用（データディスク用）に、残りの1台がパリティデータ格納用（パリティディスク用）に割り当てられる。

【0029】ディスクアレ装置20（内のディスクアレコントローラ22）では、HDD210-0～210-3によって実現されるディスクアレ21のディスク領域を、図2に示すように複数のストライプ23に分割して管理する。このストライプ23のサイズは、1HDD当たり64K（キロ）バイト～256Kバイト程度に設定されるのが一般的である。ストライプ23は、少なくとも1つの論理ブロック24から構成される。この論理ブロック24は、ホスト計算機10のファイルシステム11がディスクアレ装置20（内のディスクアレ装置20）のディスク領域を管理する際の最小単位である。つまり、ディスクアレ装置20のディスク領域



は、当該ディスクアレイ装置 20 ではストライプ 23 を単位に管理され、ホスト計算機 10 では論理ブロック 24 を単位に管理される。通常、論理ブロックのサイズは 1 K バイト～8 K バイト程度である。論理ブロック 24 は、連続する複数の物理セクタブロック 25 から構成される。このセクタブロック 25 のサイズは 512 バイトであるのが一般的である。

【0030】図 3 は、ホスト計算機 10 内に保持されるディスク資源管理テーブル 12 とディスクアレイコントローラ 22 のメモリ 222 内のディスク資源管理情報領域 222b に格納されるディスク資源管理テーブル 120 のデータ構造例を示す。

【0031】ディスク資源管理テーブル 12 の各エントリの並び順で決まるエントリ番号は、そのまま論理ブロック番号を表すようになっている。このテーブル 12 の各エントリには、そのエントリに固有の論理ブロック番号で表される論理ブロックがファイルシステム 11 により使用されているか否かを示すフラグが設定されている。なお、ディスク資源管理テーブル 12 の各エントリに、論理ブロック番号と上記フラグとの対が設定されるものであっても構わない。

【0032】一方、ディスク資源管理テーブル 120 の各エントリの並び順で決まるエントリ番号は、そのままストライプ番号を表すようになっている。このテーブル 120 の各エントリには、そのエントリに固有のストライプ番号で表されるストライプ（に含まれる論理ブロックの少なくとも 1 つ）がファイルシステム 11 により使用されているか否かを示すフラグが設定されている。なお、ディスク資源管理テーブル 120 の各エントリに、ストライプ番号と上記フラグとの対が設定されるものであっても構わない。

【0033】次に、図 1 の構成の計算機システムにおける動作を、(1) ホスト計算機 10 からのディスク資源管理情報リスト送信時の処理、(2) ホスト計算機 10 からのデータ書き込み要求発行時の処理、(3) ディスクアレイ装置 20 におけるデータ修復処理、(4) ディスクアレイ装置 20 における HDD メディア障害検査処理を例に順に説明する。

【0034】(1) ホスト計算機 10 からのディスク資源管理情報リスト送信時の処理

まず、ホスト計算機 10 からのディスク資源管理情報リスト送信時の処理について、図 4 のフローチャートを参照して説明する。

【0035】ホスト計算機 10 は、当該計算機 10（の記憶装置）にインストールされている専用ソフトウェア 13 に従い、予め定められたタイミングで、その時点に当該計算機 10（の記憶装置）に保持されているディスク資源管理テーブル 12 をファイルシステム 11 から取得する。そしてホスト計算機 10 は、このディスク資源管理テーブル 12 から図 9 に示すディスク資源管理情報

リスト 90 を作成し、当該リスト 90 をディスクアレイ装置 20 に送信する。このときホスト計算機 10 は、ディスク資源管理テーブル 12 の内容がディスク資源管理情報リスト 90 の送信中に変化するのを防止するために、ファイルの更新が発生しないように配慮することが好ましい。また、ディスク資源管理情報リスト 90 は極めて大きなサイズとなる可能性があり、その場合には当該リスト 90 の送信に長時間を要する。そこで、ディスク資源管理情報リスト 90 の送信がホスト計算機 10 の効率に影響を及ぼさないように、送信タイミングとして、ホスト計算機 10 の立ち上げ時、或いはホスト計算機 10 の負荷が少ない夜間等の一定周期を設定するとよい。

【0036】さて、ホスト計算機 10 からディスクアレイ装置 20 に送信されるディスク資源管理情報リスト 90 は、図 9 に示すように、ホスト計算機 10 内のファイルシステム 11 がディスクアレイ装置 20 のディスク領域（ディスクボリューム）を扱う際の論理ブロック 24 のサイズ 91 と、このディスク領域内のすべての論理ブロック 24 のうち、ファイルシステム 11 により使用されている論理ブロック 24 のブロック番号（論理ブロック番号）92、92…の集合とから構成される。このように、ディスク資源管理情報リスト 90 内に、ファイルシステム 11 により使用されていない論理ブロック 24 の情報（論理ブロック番号）が含まれていないのは、当該リスト 90 のサイズを小さくすることで、当該リスト 90 をホスト計算機 10 からディスクアレイ装置 20 に送信するのに要する時間を短縮するためである。通常、ディスクアレイ装置 20 のディスク領域のうち、ファイルシステム 11 によって使用されている領域の占める割合は少ない。このような場合、ファイルシステム 11 により使用されていない論理ブロック 24 の情報をディスク資源管理情報リスト 90 に含まないことは、当該リスト 90 の送信時間を短縮するのに特に効果がある。なお、ディスク資源管理テーブル 12 自体をホスト計算機 10 からディスクアレイ装置 20 に送信するようにしても構わない。また、ディスク資源管理情報リスト 90 またはディスク資源管理テーブル 120 のうちデータ量の少ない方を送信するようにしても構わない。この場合、リスト 90 またはテーブル 120 のいずれの送信であるかを示す情報を付加して送信するとよい。

【0037】ディスクアレイ装置 20 内のディスクアレイコントローラ 22（に設けられたマイクロプロセッサ 221）は、ホスト計算機 10 からディスク資源管理情報リスト 90 が送信されると、当該リスト 90 を受信する（ステップ S1）。するとディスクアレイコントローラ 22 は、ディスク資源管理情報リスト 90 に含まれているすべての論理ブロック番号をもとに、ファイルシステム 11 により使用されていない論理ブロックの論理ブロック番号を判別し、ディスクアレイ装置 20 のディス

ク領域内のすべての論理ブロックについて、そのブロックを示す論理ブロック番号の例えば昇順に、そのブロックが使用されているか否かを示すエントリが配置された、図3に示すディスク資源管理テーブル12を復元する(ステップS2)。

【0038】次にディスクアレイコントローラ22は、ホスト計算機10により管理される論理ブロック番号とディスクアレイ装置20により管理されるストライプ番号との対応付けを行う(ステップS3)。この対応付けは次のように行われる。

【0039】まずディスクアレイコントローラ22は、「ストライプ当たりの論理ブロック数」を、自身が管理している「ストライプのサイズ」と、ホスト計算機10から送信されたディスク資源管理情報リスト90に含まれている「論理ブロックのサイズ」91とから、「ストライプ当たりの論理ブロック数」=「ストライプのサイズ」/「論理ブロックのサイズ」により算出する。

【0040】次にディスクアレイコントローラ22は、すべての「論理ブロック番号」について、その「論理ブロック番号」と「ストライプ当たりの論理ブロック数」とから、「論理ブロック番号」の示す論理ブロック24が含まれているストライプ23を示す「ストライプ番号」を、

「ストライプ番号」= {「論理ブロック番号」/「ストライプ当たりの論理ブロック数」} の整数部により算出する。例えば、「ストライプ当たりの論理ブロック数」を「4」とすると、「論理ブロック番号」が「0」～「3」の論理ブロックを含むストライプの「ストライプ番号」はいずれも「0」である。この結果、ホスト計算機10により管理される論理ブロック番号とディスクアレイ装置20により管理されるストライプ番号との対応付けが行われたことになる。

【0041】ディスクアレイコントローラ22は、論理ブロック番号とストライプ番号との対応付けを行うと、その対応付けの結果と復元されたディスク資源管理テーブル12とから、ディスクアレイ装置20内のすべてのストライプについて、そのストライプを示すストライプ番号の例えば昇順に、そのストライプがファイルシステム11により使用されているか否かを示すエントリが配置された、図3に示すディスク資源管理テーブル120を作成する(ステップS4)。ここでは、ファイルシステム11によって使用されている論理ブロックを1つでも含むストライプは、ファイルシステム11により使用されていると判定されて、対応するエントリに使用中を示すフラグが設定される。これに対し、ファイルシステム11によって使用されている論理ブロックを含まないストライプは、ファイルシステム11によって使用されていないと判定されて、対応するエントリに未使用(不使用)を示すフラグが設定される。

【0042】ディスクアレイコントローラ22はディスク資源管理テーブル120を作成すると、当該テーブル120をメモリ222内のディスク資源管理テーブル領域222bに上書きコピーする(ステップS5)。

【0043】上記のように、論理ブロック番号とストライプ番号との対応付けをディスクアレイ装置20で行う場合、ホスト計算機10ではディスクアレイ装置20に固有のストライプのサイズを考慮する必要がない。但し、ディスクアレイ装置20では、ホスト計算機10に固有の論理ブロックのサイズを考慮する必要がある。

【0044】これに対し、図9のディスク資源管理情報リスト90に代えて、図10のデータ構造のディスク資源管理情報リスト100を用いるならば、ディスクアレイ装置20では、ホスト計算機10に固有の論理ブロックのサイズを考慮する必要がない。この図10の構造のディスク資源管理情報リスト100は、ファイルシステム11により使用されている論理ブロック24が含まれるストライプを示すストライプ番号101、101…の集合から構成される。但し、図10の構造のディスク資源管理情報リスト100をホスト計算機10で用意するには、当該ホスト計算機10がディスクアレイ装置20内のディスクアレイコントローラ22から予めストライプのサイズを取得し、論理ブロック番号とストライプ番号との対応付けを、当該ホスト計算機10にインストールされている専用ソフトウェア13に従って実行する必要がある。

【0045】(2)ホスト計算機10からのデータ書き込み要求発行時の処理

次に、メモリ222内のディスク資源管理テーブル領域222bに、図3に示すディスク資源管理テーブル120が格納されている状態で、ホスト計算機10からディスクアレイ装置20内のディスクアレイコントローラ22に対してデータの書き込み要求が発行された場合の処理について、図5のフローチャートを参照して説明する。

【0046】まずディスクアレイコントローラ22(内のマイクロプロセッサ221)は、ホスト計算機10(内のファイルシステム11)からディスクアレイ装置20に対してデータ書き込み要求が発行されると、当該要求を受信する(ステップS11)。この要求には、アドレス(開始アドレス)とサイズとが含まれている。

【0047】次にディスクアレイコントローラ22は、受信したデータ書き込み要求に含まれている開始アドレスとサイズとから、書き込み対象となるストライプを示すストライプ番号を算出する(ステップS12)。

【0048】次にディスクアレイコントローラ22は、ディスク資源管理テーブル領域222bに格納されているディスク資源管理テーブル120内のエントリのうち、ステップS12で算出したストライプ番号で指定されるエントリを参照することにより、当該ストライプ番

号の示す書き込み対象ストライプがファイルシステム11にて既に使用されているか否かを判定する(ステップS13)。もし、書き込み対象ストライプがそれまで使用されていなかった場合、ディスクアレイコントローラ22はステップS13で参照したディスク資源管理テーブル120内のエントリの内容(フラグの状態)を、未使用から使用中を示すように更新する(ステップS14)。このように、ホスト計算機10からのデータ書き込み要求で指定されるデータ書き込みにより、当該要求で指定される論理ブロックを含むストライプの状態が未使用から使用中に変化すると判定された場合には、当該ストライプに対応するディスク資源管理テーブル120内のエントリの内容が使用中を示すように更新される。

【0049】ディスクアレイコントローラ22はステップS14を実行すると、ディスクアレイ21に対してホスト計算機10からのデータ書き込み要求で指定されたデータ書き込みを行う(ステップS15)。

【0050】またディスクアレイコントローラ22は、ステップS12で算出したストライプ番号の示す書き込み対象ストライプがファイルシステム11にて既に使用されている場合には(ステップS13)、ステップS14をスキップしてステップS15に進み、ホスト計算機10からのデータ書き込み要求で指定されたデータ書き込みを行う。

【0051】(3) ディスクアレイ装置20におけるデータ修復処理

次に、ディスクアレイ装置20におけるデータ修復処理について、図6のフローチャートを参照して説明する。

【0052】今、ディスクアレイ21内のHDD210-0~210-3のうちHDD210-3が故障したために、その故障したHDD(旧HDD)210-3を新たなHDD(新HDD)210-3に交換して、旧HDD210-3内のデータを新HDD210-3に修復するものとする。ここでは便宜的に、新HDD、つまり修復先となるHDDにも、故障した旧HDD210-3と同一符号“210-3”を付してある。なお、修復先となるHDDがディスクアレイコントローラ22に予め接続されているスベアHDDであっても構わない。

【0053】ディスクアレイコントローラ22(内のマイクロプロセッサ221)はHDD210-3が故障した場合、その故障したHDD(旧HDD)210-3内のデータを、新HDD210-3に修復する処理を、ストライプ番号nが0の先頭のストライプから順番に次のように実行する(ステップS21)。

【0054】まずディスクアレイコントローラ22は、ストライプ番号n(nの初期値は0)のストライプ23のデータ修復のために、そのストライプ番号nで指定されるディスク資源管理テーブル120内のエントリを参照して、そのストライプ番号nの示すストライプ23(に含まれている論理ブロックの少なくとも1つ)がフ

ァイルシステム11により使用されているか否かを判定する(ステップS22)。

【0055】もし、ストライプ番号nの示すストライプ23が使用されているならば、ディスクアレイコントローラ22は、従来から知られているRAIDの機能に従う通常のデータ修復処理を以下に述べる手順で図11に示すように行う。

【0056】まず、ディスクアレイコントローラ22は、修復するストライプ23について、正常なすべてのHDD210-0~210-2からのデータ読み出し111を行う(ステップS23)。次にディスクアレイコントローラ22は、読み出したデータを使用してRAIDの機能により、つまり排他的論理和演算112により、その演算結果として修復されたデータを取得する(ステップS24)。そしてディスクアレイコントローラ22は、取得したデータをストライプ23に含まれる新HDD210-3内の領域に書き込む動作113を実行する(ステップS25)。これにより旧HDD210-3内のデータが新HDD210-3に修復される。

【0057】これに対し、ストライプ番号nの示すストライプ23がファイルシステム11により使用されていないならば、ディスクアレイコントローラ22は当該ストライプ23内には修復すべき有効なデータが格納されていないものと判断する。この場合、ディスクアレイコントローラ22は図12に示す動作を行う。即ちディスクアレイコントローラ22は、ストライプ23に含まれる正常なすべてのHDD210-0~210-2内の各領域に対して予め定められた固定データを書き込む動作211を実行すると共に、その固定データの排他的論理和値をストライプ23に含まれる新HDD210-3内の領域に書き込む動作212を実行する(ステップS26)。ここで、固定データの排他的論理和値は、実際に排他的論理和演算を行うことにより取得されるものでも、予め定められた固定値であっても構わない。つまり、ステップS26では、ファイルシステム11により使用されていないストライプ23への固定データの書き込みが行われる。このステップS26の動作は、正常なHDD210-0~210-2からの読み出しを必要としないため、ファイルシステム11により使用されているストライプ23のデータを修復する場合(ステップS23~S25)に比べて、短時間で実行できる。また、HDD210-0~210-2からの読み出しが行われないことにより、HDDの多重障害となる危険性が大幅に低下する。

【0058】ディスクアレイコントローラ22はステップS25またはS26を終了すると、ストライプ番号nを1だけインクリメントし(ステップS27)、そのインクリメント後のストライプ番号nが最終ストライプ番号を越えたか否かにより、最終ストライプまで修復処理を終了したか否かを判定する(ステップS28)。ディスクアレイコントローラ22は、ステップS28で未終

了を判定したならば、ステップS21以降の動作を再度実行し、終了を判定したならば一連のデータ修復処理を終了する。

【0059】なお、以上に述べた故障したHDD内のデータを修復する処理では、ファイルシステム11により使用されていないストライプについては、当該ストライプへの固定データの書き込み（ステップS26）が行われるものとして説明した。しかし、ファイルシステム11により使用されていないストライプ内には修復すべき有効なデータが格納されていないことから、図6のフローチャートにおいて破線60で示すように、このステップS26の動作（固定データによるストライプの修復動作）をスキップするようにしても構わない。

【0060】しかし、ステップS26をスキップすると、ファイルシステム11により使用されていないストライプについては、RAIDレベル5におけるデータと冗長データ（パリティデータ）との整合性が得られなくなる。つまり、データに対して正しいパリティデータが生成されていない状態となる。したがって、ステップS26の動作をスキップする方法を適用する場合、ホスト計算機10からのデータ書き込み要求を、図7のフローチャートに示す手順に従って図13に示すように処理する必要がある。

【0061】まずディスクアレイコントローラ22は、ホスト計算機10からディスクアレイ装置20に対してデータ書き込み要求が発行された場合、当該要求で指定されたデータの書き込み対象となるストライプ23のストライプ番号を算出し、そのストライプ23がファイルシステム11により使用されているか否かをディスク資源管理テーブル120に基づいて判定する（ステップS31～S33）。ここまでは図5のフローチャートのステップS11～S13と同様である。

【0062】もし、データの書き込み対象となるストライプ23がファイルシステム11により使用されていない場合、ディスクアレイコントローラ22は当該ストライプ23はRAIDレベル5によるパリティデータの整合性が得られていないものと判断する。この場合、まずディスクアレイコントローラ22は、ホスト計算機10からのデータ書き込み要求で指定された書き込みデータ（新規書き込みデータ）131と予め定められた、“ディスクアレイ21を構成するHDDの数-2”個のHDD用の固定データ133との排他的論理和値135を正しいパリティデータ（冗長データ）として取得する（ステップS34）。ここで、固定データ133を全ビットが“0”のデータとするならば、排他的論理和値135は書き込みデータ131に一致する。この場合、書き込みデータ131を排他的論理和値（冗長データ）135とすることができ、排他的論理和演算を必要としない。

【0063】次にディスクアレイコントローラ22は、

データの書き込み対象となるストライプ23に含まれるすべてのHDD210-0～210-3内の領域に、各HDD毎に、書き込みデータ131、固定データ133、または排他的論理和値（冗長データ）135を書き込む（ステップS35）。ここでは、ホスト計算機10からのデータ書き込み要求で指定されたデータの書き込み先HDDがHDD210-0であり、ストライプ23において冗長データが格納されているHDDがHDD210-3であるものとする、HDD210-0に対する書き込みデータ131の書き込み132と、HDD210-1及び210-2に対する固定データ133の書き込み134と、HDD210-3に対する排他的論理和値（冗長データ）135の書き込み136とが、それぞれ行われる。これにより、故障したHDDのデータ修復の際に修復をスキップしたストライプについてもデータの冗長性を保証できる。

【0064】ディスクアレイコントローラ22はステップS35を終了すると、ステップS33で参照したディスク資源管理テーブル120内のエントリの内容、つまりデータの書き込み対象となったストライプ23の使用の有無を示すディスク資源管理テーブル120内のエントリの内容を、未使用から使用中を示すように更新する（ステップS36）。

【0065】なお、データの書き込み対象となるストライプ23がファイルシステム11により使用されている場合には、ディスクアレイコントローラ22は通常のRAID手法によるデータ書き込みを行う（ステップS37）。ここでは、データ書き込み要求で指定されたデータ（新データ）と、当該新データの格納先となるHDD内領域に格納されているデータ（旧データ）と、同じストライプ23に含まれている別のHDD内領域に格納されているパリティデータ（旧パリティデータ）との間の排他的論理和演算を行うことで、新パリティデータ（新冗長データ）を生成し、当該新パリティデータで旧パリティデータを更新する。

【0066】（4）ディスクアレイ装置20におけるHDDメディア障害検査処理  
次に、ディスクアレイ装置20におけるHDDメディア障害検査処理について、図8のフローチャートを参照して説明する。

【0067】ディスクアレイコントローラ22は、HDDメディア検査を例えば当該コントローラ22の有するパトロール機能により周期的に実行する。ここではディスクアレイコントローラ22は、HDDメディア検査を、ストライプ番号nが0の先頭のストライプから順番に次のように実行する（ステップS41）。

【0068】まずディスクアレイコントローラ22は、ストライプ番号n（nの初期値は0）のストライプ23のHDDメディア検査のために、そのストライプ番号nで指定されるディスク資源管理テーブル120内のエン

トリを参照して、そのストライプ番号nの示すストライプ23がファイルシステム11により使用されているか否かを判定する（ステップS42）。

【0069】もし、ストライプ番号nの示すストライプ23が使用されているならば、ディスクアレイコントローラ22は、当該ストライプについて、すべてのHDD210-0～210-3からのデータ読み出しを行う（ステップS43）。

【0070】次にディスクアレイコントローラ22は、HDD210-0～210-3からのデータ読み出し結果をチェックして、読み出しに成功したか否かを判定する（ステップS44）。

【0071】もし、HDD210-0～210-3のいずれかからのデータ読み出しに失敗したならば、ディスクアレイコントローラ22は、その失敗したセクタブロック、つまりメディア障害が検出された不良セクタブロックを、同じHDD内の別のセクタ（交代セクタ）に代替する代替処理を行う（ステップS45）。例えば、図14に示すように、HDD210-1内のセクタブロック141が不良セクタブロックとして検出された場合であれば、当該セクタブロック141を同じHDD210-1内の任意の交代セクタ142に代替する代替処理143が行われる。

【0072】次にディスクアレイコントローラ22は、RAIDの機能を使用して、不良セクタブロック141の修復されたデータを算出し、そのデータを交代セクタ142に書き込む動作144を行う（ステップS46）。そしてディスクアレイコントローラ22はステップS47に進む。

【0073】これに対し、ストライプ番号nの示すストライプ23についてHDD210-0～210-3からのデータ読み出しに成功したならば、ディスクアレイコントローラ22はステップS45、S46をスキップしてステップS47に進む。

【0074】また、ストライプ番号nの示すストライプ23がファイルシステム11によって使用されていないならば、ディスクアレイコントローラ22は当該ストライプ23の検査をせずに、ステップS43～S46をスキップしてステップS47に進む。

【0075】ディスクアレイコントローラ22は、ステップS47においてストライプ番号nを1だけインクリメントし、そのインクリメント後のストライプ番号nが最終ストライプ番号を越えるまで（ステップS48）、ステップS41以降の動作を繰り返す。

【0076】以上に述べたHDDメディア障害検査処理では、ファイルシステム11により使用されているストライプのみ、HDDからのデータ読み出しによる検査（メディア検査）を実行する場合について説明したが、これに限るものではない。例えば図15のフローチャートに示すように、ディスクアレイ装置20のディスク領

域のすべてのストライプについて、その使用の有無に無関係にHDDからのデータ読み出しによる検査を実行するようにしてもよい（ステップS52、S53）。ここでは、データ読み出しによる検査で、メディア障害となったセクタブロック（不良セクタブロック）が検出された場合に、ディスク資源管理テーブル120を参照して、そのセクタブロックを含むストライプがファイルシステム11により使用されているか否かを判定する（ステップS54）。

【0077】もし、上記ストライプがファイルシステム11により使用されていない場合、そのストライプ内のデータを保持する必要性はない。そこで、この場合には、RAIDの冗長性の有無に拘わらずに、図12に示すように不良セクタブロック（141）の代替処理（143）を行った後に（ステップS55）、交代セクタ（142）に（修復データではなくて）予め定められた固定データの書き込み（144）を行って修復する（ステップS56）。このステップS56では、上記ステップS26におけるストライプに対するのと同様に、交代セクタに対応する他の正常なHDD内のセクタに固定データが書き込まれる。

【0078】一方、上記ストライプがファイルシステム11により使用されている場合、他のHDDにて故障などが発生しておらず（ステップS57）、したがってデータの冗長性が確保されている場合は、不良セクタブロック（141）を交代セクタ（142）に代替する処理（143）を行った後（ステップS58）、RAIDの機能を使用して、不良セクタブロック（141）のデータを修復し、その修復データの交代セクタ（142）への書き込み（144）を行う（ステップS59）。

【0079】これに対し、他のHDDが故障しているためにデータの冗長性がなくなっている場合は（ステップS57）、不良セクタブロック（141）のデータをRAIDの機能により修復することができない。この場合、上記ステップS58、S59をスキップし、不良セクタブロック（141）をそのまま放置する。

【0080】なお、以上の実施形態では、ディスクアレイ装置20がRAID5レベルで用いられるものとして説明したが、本発明は、RAID3レベルなど、他のRAIDレベルで用いられるディスクアレイ装置にも、データ修復の方法は異なるものの、RAID5レベルの場合と同様に適用できる。

【0081】なお、本発明は、上記実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。更に、上記実施形態には種々の段階の発明が含まれており、開示される複数の構成要件における適宜な組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件から幾つかの構成要件が削除されても、発明が解決しようとする課題の欄で述べた課題が解決でき、発明

の効果の欄で述べられている効果の少なくとも1つが得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

【0082】

【発明の効果】以上詳述したように本発明によれば、ディスクアレイのディスク領域のうちファイルシステムにて実際に使用されている領域を判定し、このファイルシステムにて実際に使用されている領域に絞ってディスクドライブのメディア障害検査を行うようにしたので、処理時間を短縮することができ、その結果、メディア障害の早期検出が可能となり、ディスクアレイ装置の信頼性を向上することができる。

【0083】また本発明によれば、ホスト計算機からデータ書き込み要求を受け取った場合、当該要求で指定されたデータ書き込み先が属するストライプが、ファイルシステムにより使用されている論理ブロックを含む第1のストライプであるか或いはファイルシステムにより使用されている論理ブロックを含まない第2のストライプであるかを、ディスクアレイ装置内に保持されているディスク資源管理情報（第2のディスク資源管理情報）に基づいて判定し、第2のストライプであるならば、当該データ書き込み先が属するストライプが第1のストライプであることを示すように第2のディスク資源管理情報を更新するようにしたので、当該第2のディスク資源管理情報によりストライプの最新の使用状況を表すことができ、ファイルの更新の度に最新の第1のディスク資源管理情報をホスト計算機から取得しなくても済む。

【図面の簡単な説明】

【図1】本発明の一実施形態に係るディスクアレイ装置20を備えた計算機システムの構成を示すブロック図。

【図2】ディスクアレイ21のディスク領域を管理するのに用いられるストライプ、論理ブロック及びセクタブロックの関係を説明するための図。

【図3】ホスト計算機10内に保持されるディスク資源管理テーブル12とディスクアレイコントローラ22のメモリ222内のディスク資源管理情報領域222bに格納されるディスク資源管理テーブル120のデータ構造例を示す図。

【図4】ディスク資源管理情報リスト送信時の処理手順を示すフローチャート。

【図5】ホスト計算機10からのデータ書き込み要求発

行時の処理手順を示すフローチャート。

【図6】ディスクアレイ装置20におけるデータ修復処理手順を示すフローチャート。

【図7】ホスト計算機10からのデータ書き込み要求発行時の処理手順の変形例を示すフローチャート。

【図8】ディスクアレイ装置20におけるHDDメディア障害検査の処理手順を示すフローチャート。

【図9】ホスト計算機10からディスクアレイ装置20に送信されるディスク資源管理情報リストのデータ構造例を示す図。

【図10】上記ディスク資源管理情報リストの変形例を示す図。

【図11】RAIDレベル5における通常のデータ修復処理を説明するための図。

【図12】使用されていないストライプへの固定データ書き込み時の動作を説明するための図。

【図13】使用されていないストライプへの新規データ書き込み時の動作を説明するための図。

【図14】HDDメディア障害検査で不良セクタブロックが検出された場合のデータ修復動作を説明するための図。

【図15】図8に示したHDDメディア障害検査の処理手順の変形例を示すフローチャート

【符号の説明】

10…ホスト計算機

11…ファイルシステム

12…ディスク資源管理テーブル

13…専用ソフトウェア

20…ディスクアレイ装置

21…ディスクアレイ

22…ディスクアレイコントローラ

23…ストライプ

24…論理ブロック

25…セクタブロック

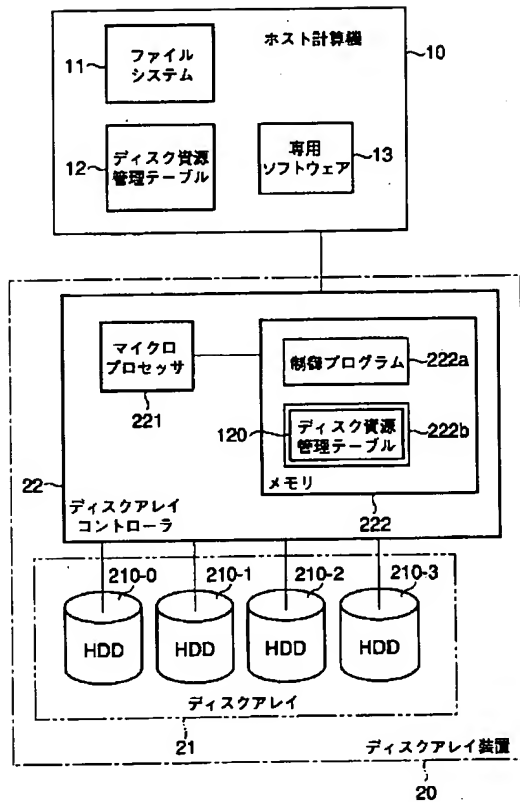
90、100…ディスク資源管理情報リスト（第1のディスク資源管理情報）

120…ディスク資源管理テーブル（第2のディスク資源管理情報）

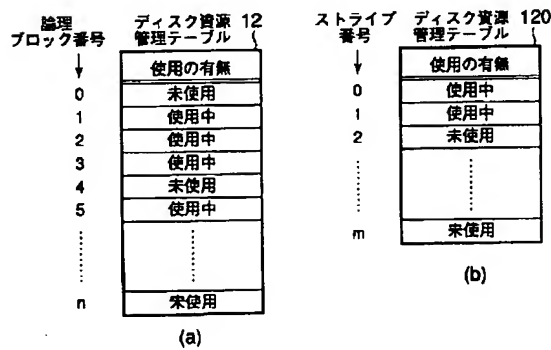
210-0～210-3…HDD（ディスクドライブ）

222b…ディスク資源管理テーブル領域

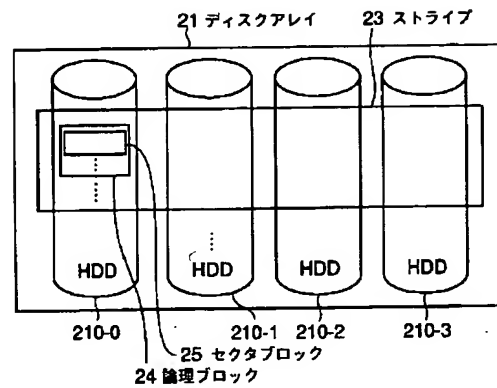
【図1】



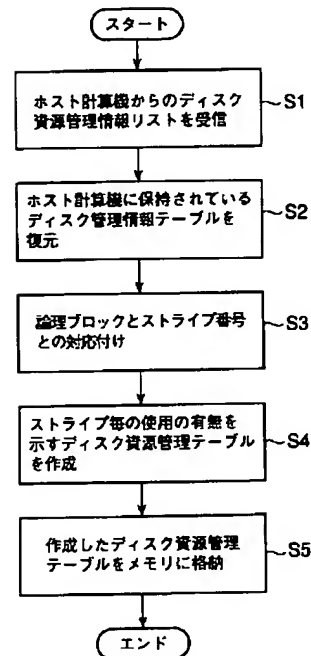
【図3】



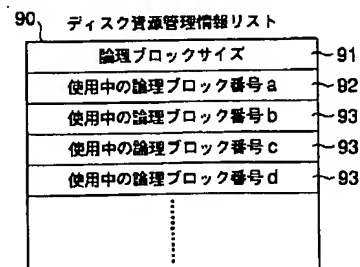
【図2】



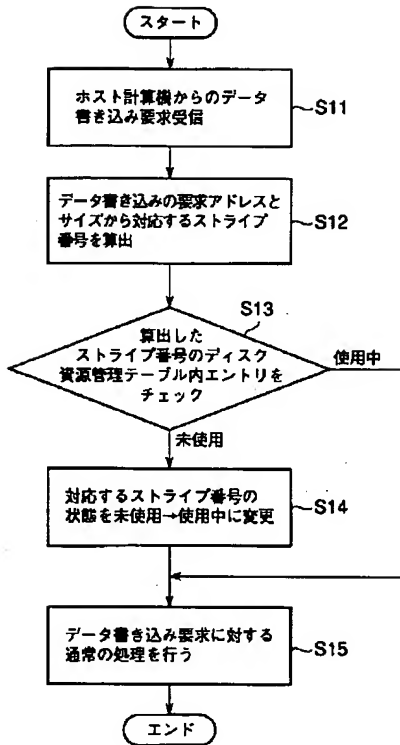
【図4】



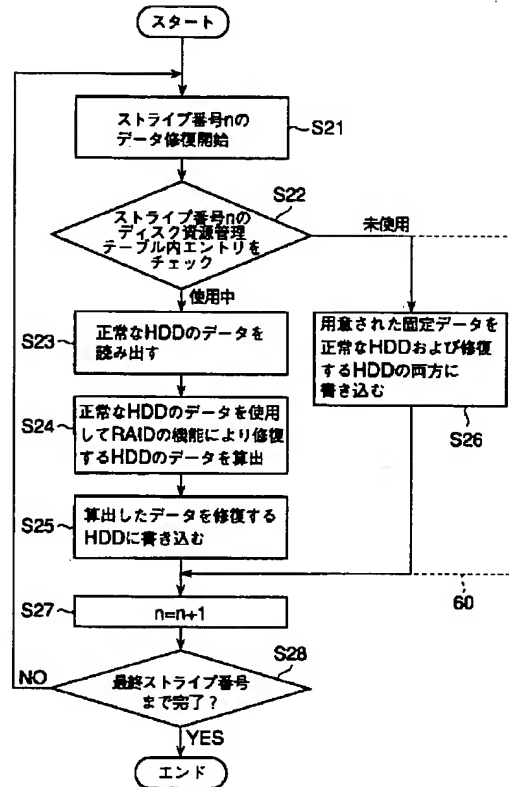
【図9】



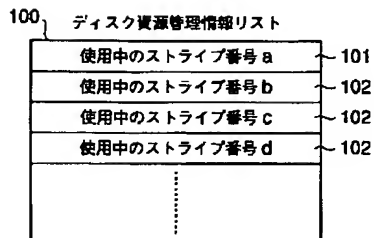
【図5】



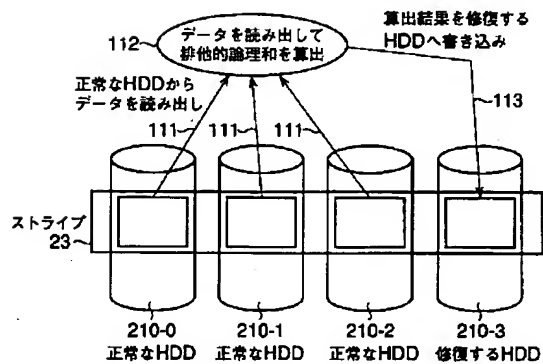
【図6】



【図10】

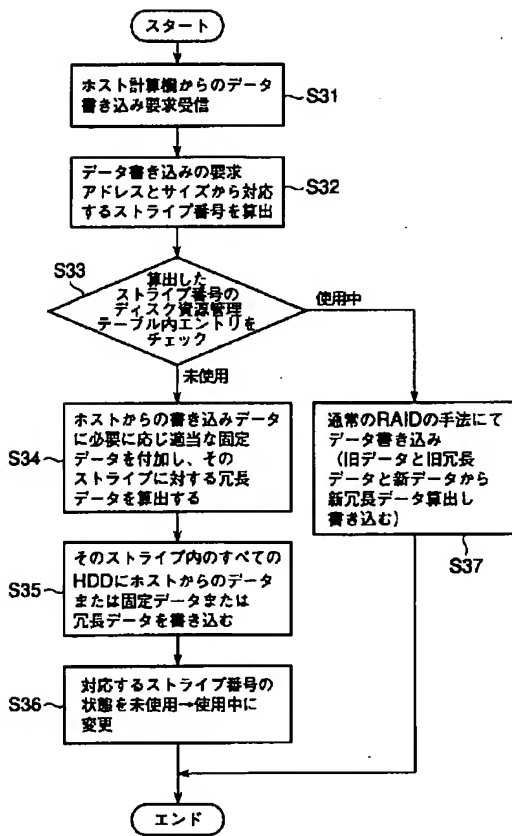


【図11】

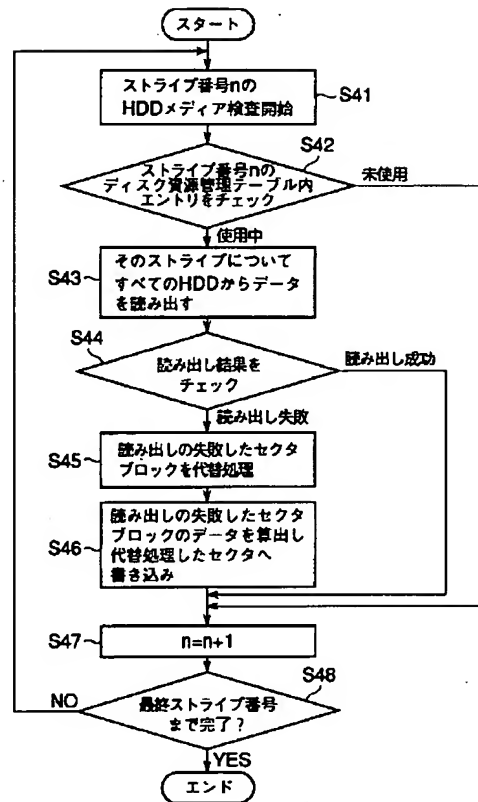




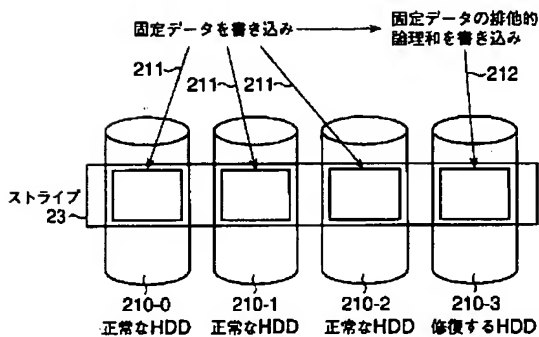
【図7】



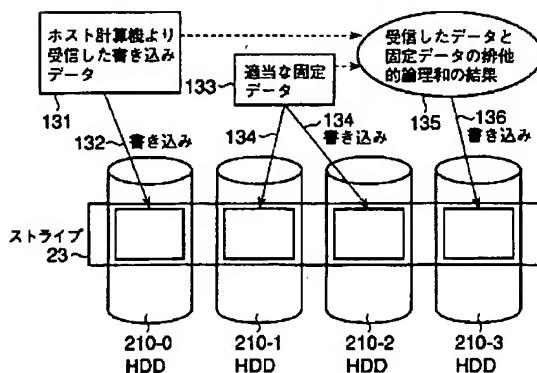
【図8】



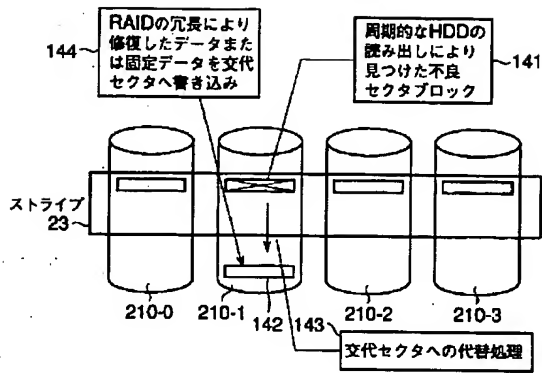
【図12】



【図13】



【図14】



【図15】

